# CAPSTONE PROJECT -SPACEX

ADITYA SAKURU

02-08-2023

# OUTLINE

- Executive Summary
- Introduction
- Data collection and data wrangling methodology
- EDA and interactive visual analytics methodology
- Predictive Analysis methodology
- Results
  - Visualization – Charts
  - Dashboard
- Conclusion

# EXECUTIVE SUMMARY

- We use Data science methodology to predict the successful launching of First stage of Falcon9
- **Collect** data using SpaceX REST API and web scraping techniques
- **Explore** data with data visualization techniques, considering the following factors: payload, launch site, flight number and yearly trend.We will achieve the features by analysing the relationship between variables and successful launching
- **Analyze** the data with SQL, calculating the following statistics: total payload, payload range for successful launches, and total # of successful and failed outcomes
- **Explore** launch site success rates and proximity to geographical markers using folium
- **Visualize** the launch sites with the most success and successful payload ranges
- **Build Models** to predict landing outcomes using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN) and predict best model based on accuracy

# INTRODUCTION

- Objective:

  To apply data science toolkit and machine learning in order to accurately predict the likelihood of the first stage rocket landing successfully, and thus determine the cost of a launch.
  •Explore the data in order to obtain more insight from the data.

# Data collection and data wrangling methodology

- First step in predictive analysis is Data collection.
- We make a get request to SpaceX API and clean requested data
- From the rocket column we learn the booster name,from the launchpad we found the name of the launch site being used, the longitude, and the latitude,from the payload we learnt the mass of the payload and the orbit that it is going to and From cores we would like to learn the outcome of the landing etc
- The data from these requests is stored in lists and are used to create a new dataframe.
- Finally we construct our dataset using the data
- In Data Wrangling, we deal with missing values.
- Finally the data set,Data_Falcon9 is created

# EDA and interactive visual analytics methodology

- **We predict if the Falco9 first stage will land successfully using Exploratory Data Analysis.**
- **We also prepare Data Feauture Engineering**
- Using EDA we observe effect of different variables like launchsite,payloadmass on launch outcome and visualise the detailed relation using scatter plot,
- We observe the effect of variables on each other like orbit vs payload,flight number and orbit type ,success rate and orbit typeand overall impact on launch outcome and visualise using scatter plot and bar graph

Github url:**IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb**

# EDA and interactive visual analytics methodology

By using data analysis and observing the relatioship between some preliminary insights about how each important variable would affect the success rate, we will select the features that will be used in success prediction in the future module.
Like Flight number,PayoadMass,Orbit,Launchsite,Flights,Gridfins,Reused,Legs and Landing Pad

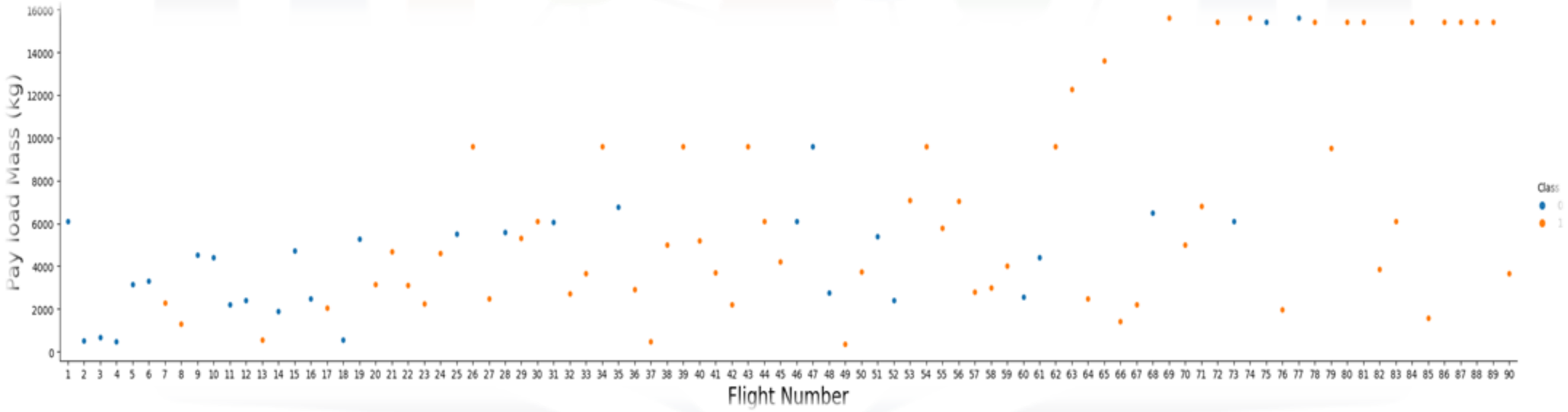
# Predictive analysis methodology

- Predict whether first stage of Falcon9 will land successfully.
- Data is Preprocessed,Split for train&test,Train data and perform Grid search on test data to obtain hyperparameters to select appropriate Model with best accuracy
- Test data is applied on Logistic Regression,Support Vector Machine,Decision classifier tree and K-nearest neighbours.

- Github url:IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

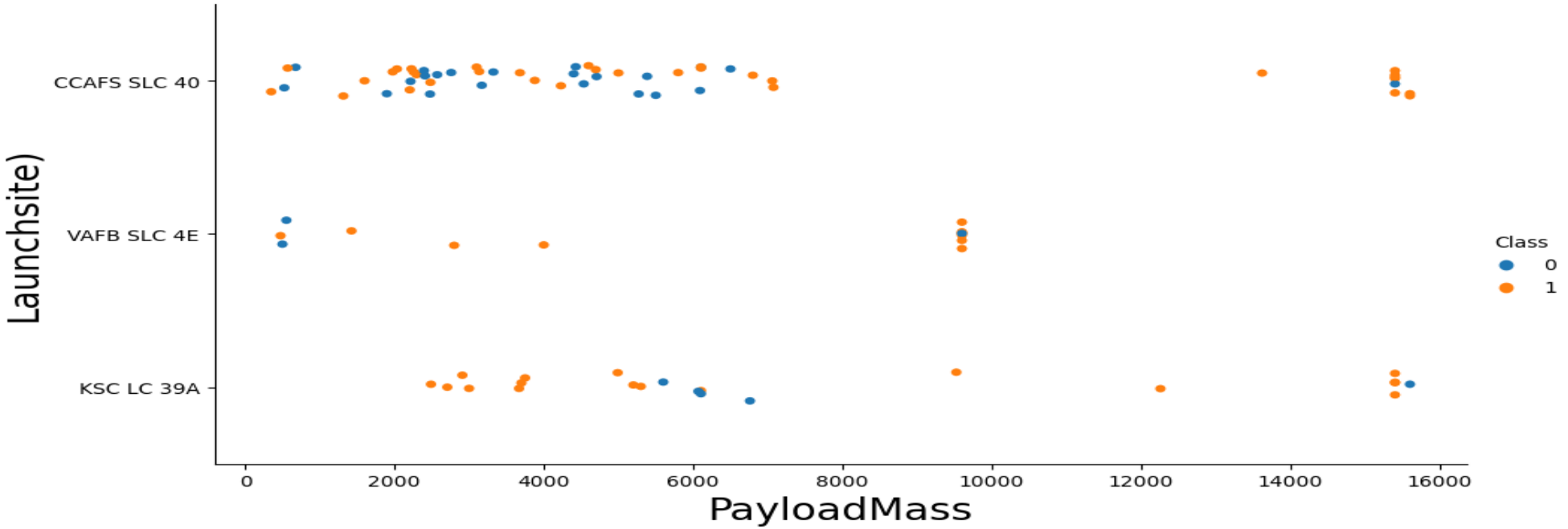
# EDA with visualization results-Flight number vs Payload mass

- We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

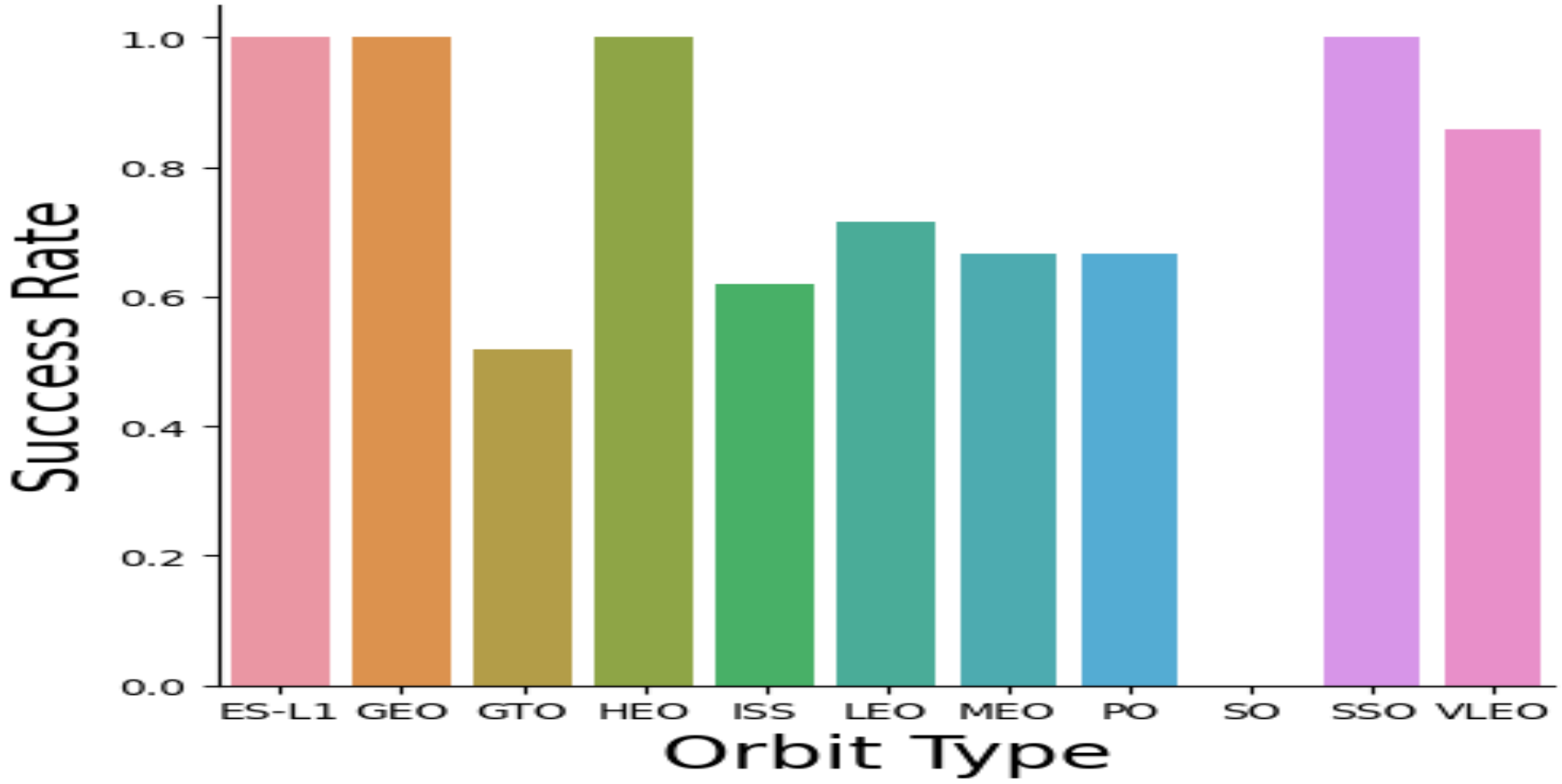


# EDA with visualization results-*relationship between Launchsite and Payloadmass*

- you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

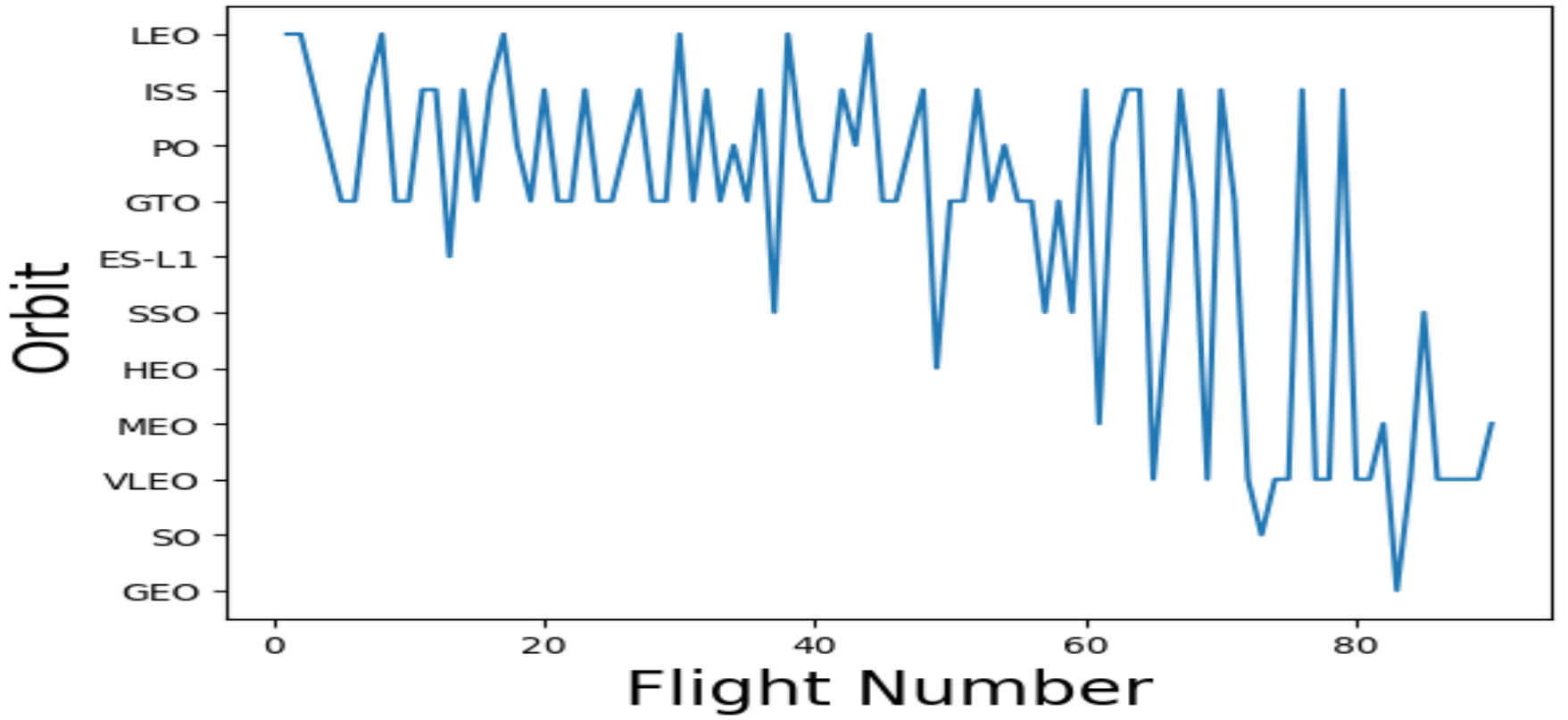


## EDA with visualization results-*Barchart for success rate of each orbit*



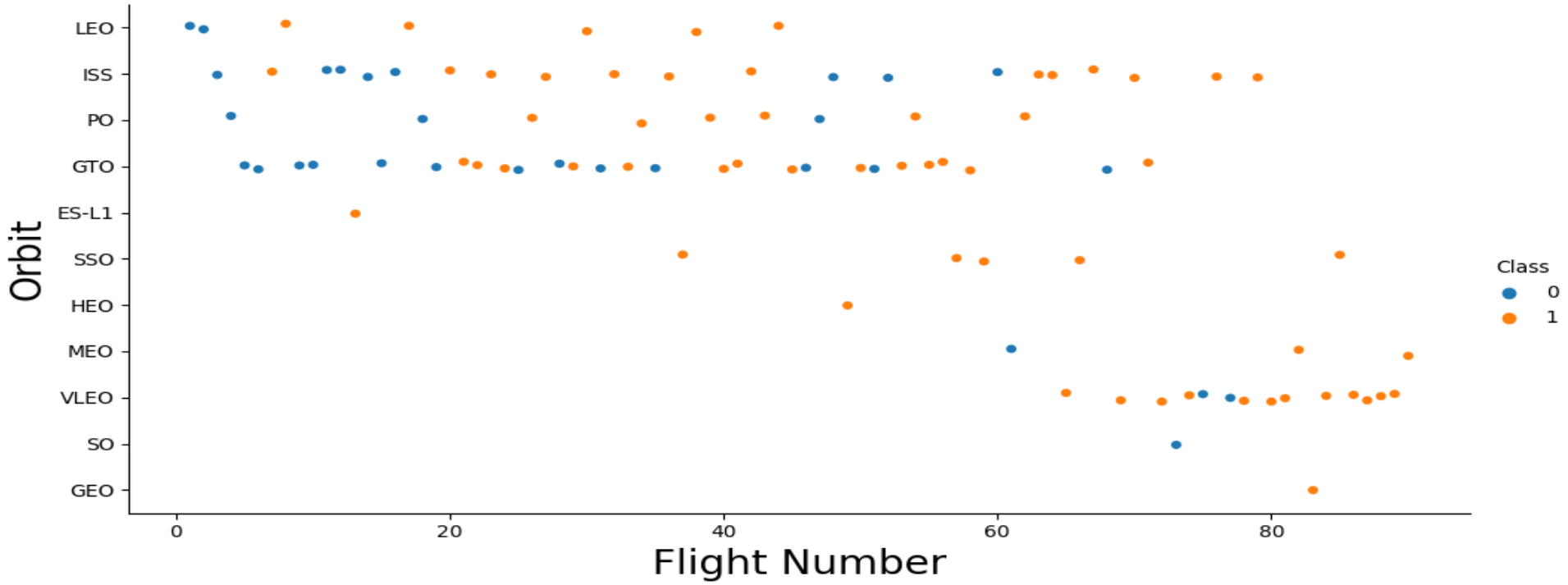
# EDA with visualization results-*relationship between orbitype and Flightnumber*

You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



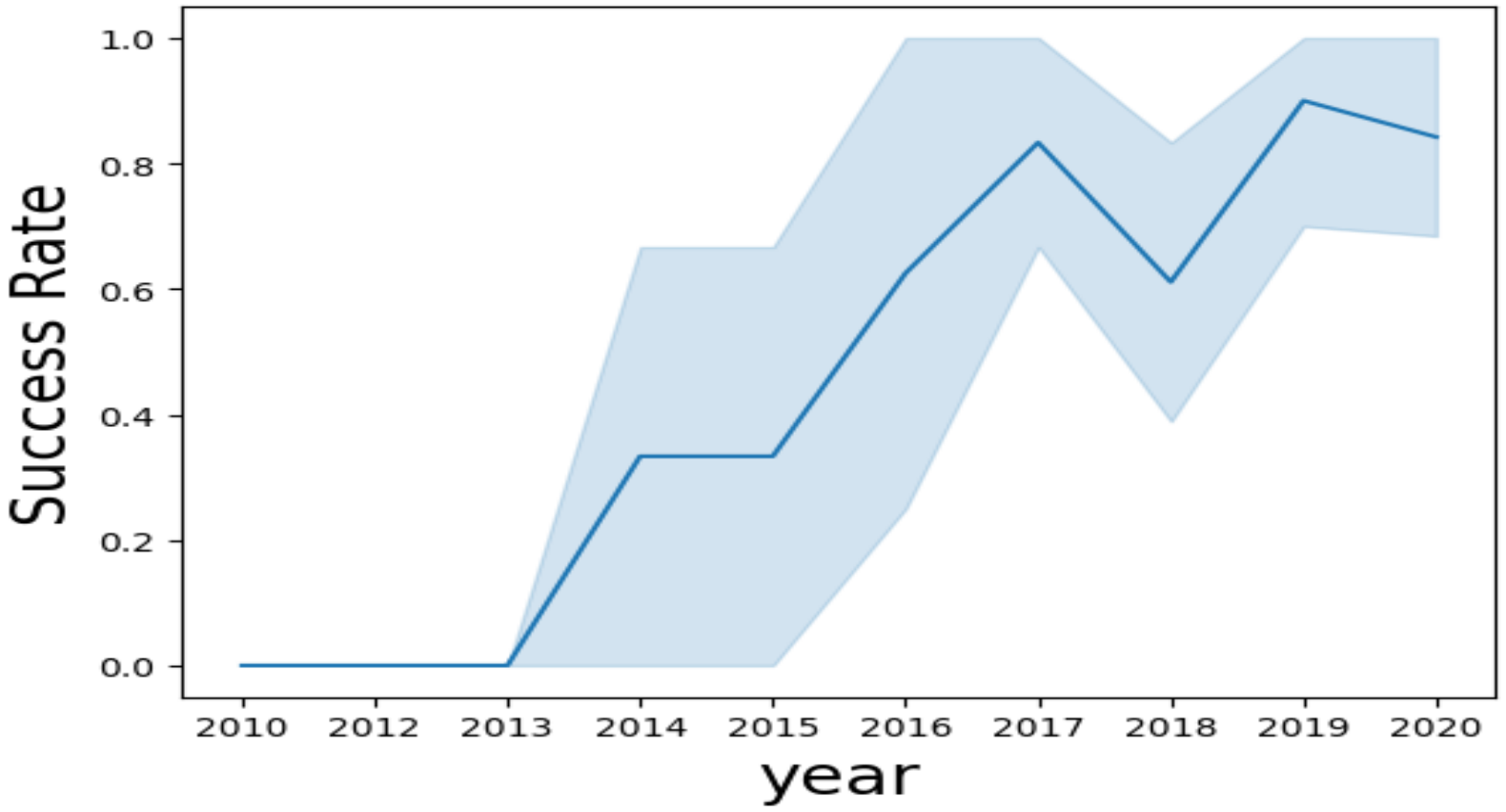
# EDA with visualization results-*relationship between payload and orbit*

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



# EDA with visualization results-*Launch success yearly trend*

Sucess rate since 2013 kept increasing till 2020



# EDA with SQL results

- Total payload mass carried by boosters launched by NASA (CRS) is 45596KG
- Average payload mass carried by booster version F9 v1.1 is 2928.4KG
- Date when the first succesful landing outcome in ground pad was achieved:

  22/12/2015; 18/07/2016; 19/02/2017; 05/01/2017;06/03/2017;14/08/2017;

  09/07/2017;15/12/2017;01/08/2018
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000:

  F9 FT B1022  F9 FT B1026  F9 FT B1021.2  F9 FT B1031.2
- Total number of successful and failure mission outcomes are 101

# EDA with SQL results

- Records which will display the month names, failure landing,outcomes in drone ship ,booster versions, launchsite for the months in year 2015:

  01/10/2015
  14/04/2015

- count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order:

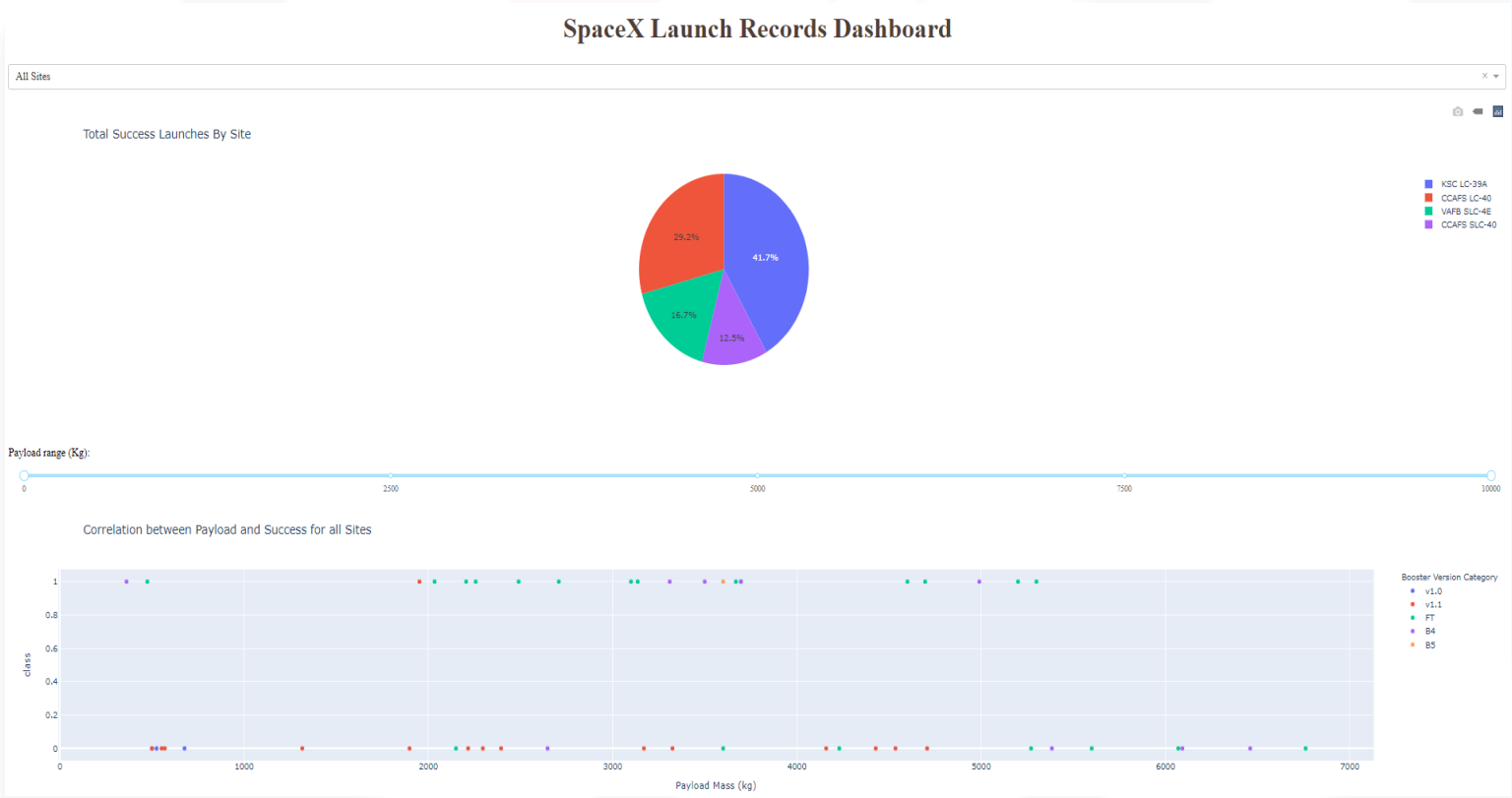
| Landing_Outcome | count_outcomes |
|---|---|
| Success | 20 |
| No attempt | 10 |
| Success (drone ship) | 8 |
| Success (ground pad) | 7 |
| Failure (drone ship) | 3 |
| Failure | 3 |
| Failure (parachute) | 2 |
| Controlled (ocean) | 2 |
| No attempt | 1 |

# Interactive map with Folium

- We can mark  all launch sites on a map,Mark the success/failed launches for each site on the map,Calculate the distances between a launch site to its proximities and  find some geographical patterns about launch sites.
- We found that launchsites are in close proxiity to coastline(0.86km) and maintain sufficient distance from railways(21.96km),highway(26.88) an City(23.23)

- Github url:**IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb**

# Plotly Dash dashboard results



# Plotly Dash dashboard results

# Predictive analysis (classification) results

We split data into training and test data and find best Hyperparameter for SVM,Logistic Regression,K Nearest Neighbour and Decision tree
ML MethodAccuracy Score (%)

- Support Vector Machine 83.333333
- Logistic Regression 83.333333
- K Nearest Neighbour 83.333333
- Decision Tree 83.333333

# Predictive analysis (classification) results

| | LogReg | SVM | Tree | KNN |
|---|---|---|---|---|
| Jaccard_Score | 0.800000 | 0.800000 | 0.615385 | 0.800000 |
| F1_Score | 0.888889 | 0.888889 | 0.761905 | 0.888889 |
| Accuracy | 0.833333 | 0.833333 | 0.833333 | 0.833333 |

Best model is DecisionTree with a score of 0.8892857142857145
Best params is : {'criterion': 'entropy', 'max_depth': 2, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'best'}

# RESULTS

- Exploratory Data Analysis:
  - Launch success has improved over time
  - KSC LC-39A has the highest success rate among landing sites
  - Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate
- Visualization / Analytics:
  - Most launch sites are near the equator, and all are close to the coast
- Predictive Analytics
  - All models performed similarly on the test set. The decision tree model slightly outperformed when looking at .best_score_

# CONCLUSION

- **Launch Success:** Increases over time
- **Model Performance:** The models performed similarly on the test set with the decision tree model slightly outperforming
- **KSC LC-39A:** Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg
- **Orbits:** ES-L1, GEO, HEO, and SSO have a 100% success rate
- **Payload Mass:** Across all launch sites, the higher the payload mass (kg), the higher the success rate